



Towards
European
Health
Data
Space

Milestone M6.2

Identification of relevant standards and data models for semantic harmonization

3 February 2022

This project has been co-funded by the European Union's 3rd Health Programme (2014-2020) under Grant Agreement no 101035467.



0 Document info

0.1 Authors

Author	Partner
Enrique Bernal-Delgado	Aragon Institute of Health Sciences (IACS)
Francisco Estupiñán-Romero	Aragon Institute of Health Sciences (IACS)
Ramón Launa-Garces	Aragon Institute of Health Sciences (IACS)
Carlos Tellería-Oriol	Aragon Institute of Health Sciences (IACS)
Thomas Engsig-Karup	Central Denmark Region (RM)
Jesper Bredmose Simonsen	Central Denmark Region (RM)

0.2 Keywords

Keywords	TEHDAS, Joint Action, Health Data, Health Data Space, Data Space, HP-JA-2020-1, data quality
-----------------	--

Accepted in Project Steering Group on 25 January 2022.

Disclaimer

The content of this deliverable represents the views of the author(s) only and is his/her/their sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use of its contents.

Copyright Notice

Copyright © 2021 TEHDAS Consortium Partners. All rights reserved. For more information on the project, please see www.tehdas.eu.

Contents

1 Executive summary	3
2 Context.....	3
2.1 TEHDAS Work Package 6: Excellence in Data Quality.....	3
2.2 Task 6.2: Develop the EHDS Semantic Interoperability Framework (EHDS-SIF).....	3
3 Methods	5
4 List of relevant standards for the harmonization of the semantic and syntactic interoperability in the EHDS2.....	7
5 Next steps	9

1 Executive summary

This milestone document contains a piece of work related to the overall work in the TEHDAS Project Work Package 6: Excellence in data quality which aims to provide solutions for the trustworthy secondary use of health and health care data with a view to promoting the digital transformation of European health systems.

This document lists a number of interoperability standards on data discoverability (at data source and variable levels) and on standards for the development of common data models, and describes some basic features: typology of interest, utility and domain/s. This list is the basis for the work to come on aiming the description of their actual use, challenges in their implementation, issues on maintenance and sustainability.

2 Context

2.1 TEHDAS Work Package 6: Excellence in Data Quality

The TEHDAS Project Plan defines the overall scope and objectives of Work Package 6 as follows:

Work Package 6 of the TEHDAS Joint Action will be providing solutions for the trustworthy secondary use of health and health care data with a view to fostering the digital transformation of the European health systems.

This overarching objective will be developed throughout two operational objectives:

- Developing the EHDS data quality assurance framework for a secondary use of real-world health data.
- Developing the EHDS Semantic Interoperability framework

2.2 Task 6.2: Develop the EHDS Semantic¹ Interoperability Framework (EHDS-SIF)

Task 6.2 will deliver on the second operational objective recommendations for the development of the EHDS SIF; thus, the output of task 6.2 will contain two types of contents: 1) A list of main relevant standards and data models for semantic harmonization in the secondary use of health data; and, 2) guidance on how to implement those standards within the context of the EHDS to achieve semantic interoperability.

As part of the former, this document lists a number of interoperability standards on data discoverability (at data source and variable levels), standards for the development of common data models, and standards for communication across nodes. For them all, it describes the basic features on: typology of interest, utility and domain. This list is the basis for the work to come aiming the description of the actual use of those standards, the challenges for their implementation, and issues on maintenance and sustainability.

¹ The conversations along the TEHDAS life and the interaction with the EHDS2 pilots working groups have recommended extending the scope of this WP6.2 to also syntactic interoperability.

3 Methods

The identification of standards builds on the following principles:

- This document elaborates **as part of the data quality framework**, thus, considering semantic (and syntactic interoperability) key to **assure comparability** across data sources, within and across data holders and, then, **obtain reliable and accurate outputs** when mobilising data in the EHDS2.
- Consistent with EHDS2 data life cycle and user's journey, this approach has to be interested both in providing recommendations on standards on **data discoverability** (at data source and variable levels), syntactic standards for **communication** across nodes, and standards that facilitate the development of semantic and syntactic **interoperable data models**.
- For both purposes, standards have to allow the development of observational studies of any kind, respond to the widest range of queries possible, from hypothesis generation to causal inference, and allow rapid-cycle analyses.
- Although semantic and syntactic interoperability operates in the whole data life-cycle, TEHDAS is **not aiming to provide guidance to data origins** (e.g., care or public health professionals, researchers, patients or citizens) on what standards they should use to collect data (such an option implies unaffordable investments for standardisation at data origins) **but to provide guidance to data institutions** (data controllers, data holders, and data permit authorities) on the type of interoperability layer suitable to respond specific queries sent out to the EHDS2. **In that sense, this piece of work is not interested in classification systems but in meta-standards, meta-data standards, ontologies, communication standards and existing interoperability layers.**

How has WP6 decided on the main standards?

The standards collected in this very milestone have been identified *via* the active participation of WP6 leaders in a number of events and working groups listed in the table below.

Date	Organiser	Content
14/1/21	PHIRI	Stakeholders' meeting => session on the requirement to build federated research infrastructures for a rapid policy response
18-19/05/21	TEHDAS	Project Forum => session on data quality and semantic interoperability
18/5/21	EMA	Data Standardisation Strategy stakeholder workshop
11/10/21	EC	Workshop Maximising investments in health research: FAIR data for a coordinated COVID-19 response
29/10/21	TEHDAS	Project Forum no 2 => session on Semantic interoperability, data quality assurance
7/12/21	EMA	EU Big Data Stakeholder Forum
2021 1st half	DG Santé	EHDS2 Pilot working groups
12/01/2022	EHealth Net	Reaction to a v.0 list in this milestone

Once a preliminary list was available, WP6 coordination team applied the aforementioned principles and released a second preliminary list for further discussion with WP6 partners. A dedicated slot presenting the approach and the list took place on December 16th. WP6 partners had to opportunity to provide feedback until December 31st.

4 List of relevant standards for the harmonization of the semantic and syntactic interoperability in the EHDS2

Acronym	Typology	Utility	Domain	Information (URL)
DCAT-AP2	Meta-data standard	Discoverability	Public reporting of data catalogues	https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en
INSPIRE	Meta-data standard	Discoverability	Data geo-allocated	https://inspire-geoportal.ec.europa.eu/#
FairSharing	Meta-standard	Discoverability	Data sources, meta-data standards, taxonomies of any health domain	https://fairsharing.org/biodbcore/?q=health+record
BBMRI/MIABIS	Meta-data standard	Discoverability	Bio samples	https://github.com/BBMRI-ERIC/miabis
Beacon	Metadata standard	Discoverability	Genomics, clinic data	https://beacon-project.io
Bioimage archive	Meta-data standard	Discoverability	Bio-Image	https://www.ebi.ac.uk/bioimage-archive/
CEDAR	Meta-data standards	Discoverability	Biomedical experiments	https://metadatacenter.org/
ECRIN	Meta-data standard	Discoverability	Randomized clinical trials	https://ecrin.org/tools/clinical-research-metadata-repository
CESSDA	Meta-data standards	Discoverability	Social data collections	https://datacatalogue.cessda.eu/
PHIRI	Meta-data standard	Discoverability	Population health data collections	https://www.healthinformationportal.eu
OMOP	Layer of interoperability	Common data model	EHR, claims data	https://ohdsi.org/omop/ ; https://athena.ohdsi.org/search-terms/terms?domain=Condition&sort=vocabulary_id&order=asc
SDTM CDISC	Layer of interoperability	Common data model	EHR, claims data	https://www.cdisc.org/
ORDO	Layer of interoperability	Taxonomy	Rare-diseases	http://www.orphadata.org/cgi-bin/index.php#ontologies ; http://www.orphadata.org/cgi-bin/index.php
SNOMED-CT	Ontology	Data provenance	HER, bio sample	https://www.snomed.org/
LOINC	Ontology	Data provenance	Lab data	https://loinc.org/ ; LOINC is now part of SNOMED-CT
ISO23494	Meta-data standard	Data provenance	Bio samples	In development here: https://www.iso.org/standard/80715.html
ISO 8000 110	Meta-data standard	Conformance messaging	Any data master file	https://www.iso.org/standard/78501.html
HL7 FHIR	Meta-data standard	Conformance messaging	EHR	https://www.hl7.org/fhir/ ; See also https://art-decor.org/mediawiki/index.php/Main_Page
DICOM	Meta-standard	Conformance messaging	Images	https://www.dicomstandard.org/

SPOR	Meta-data standard	Conformance messaging	Medicinal products (Substances, Products, Organization, Referentials)	https://www.ema.europa.eu/en/human-regulatory/research-development/data-medicines-iso-idmp-standards/substance-product-organisation-referential-spor-master-data
EDQM Standard Terms	Meta-data standard	Conformance messaging	Medicinal products (Dose forms, Routes of administration, Packaging, Units of presentation, etc.)	https://standardterms.edqm.eu/
ICF	Ontology	Common data model	Functioning	https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health
EMDN	Meta-data standard	Conformance messaging	Medical devices	https://ec.europa.eu/health/sites/default/files/md_sector/docs/md_emd_n_eudamed_nomenclature_en.pdf

5 Next steps

Milestone 1 – Mid March 2022

Main result: Scoping review on the main standards (see table above)

Survey to the coordinators
 In depth interview with the coordinators (1 to 2 hours) –
 2 thematic working groups
 Drafting table 2

Table 2 Summary of the challenges of implementation (as many tables as standards)

Challenge	List of challenges
<i>Governance requirements</i>	
<i>Technological needs: services & tools</i>	
<i>Maintenance requirements</i>	
<i>Sustainability</i>	

Milestone 2 – Mid June 2022

Main result: eliciting challenges of implementation in the EHDS (see table 2, below)

This milestone is aiming to analyse in depth challenges on governance and drivers of maintenance and sustainability

2 thematic working groups with WP6 partners
 Workshop with partners affected by the results (WP4, WP5 and WP7, specifically)

Milestone 3 – Mid October 2022

Main result – preliminary recommendations actions at different levels (LOST)

Drafting (see TOC)
 Providing WP4, WP5 and WP7 with the appropriate contents
 Public consultation to citizens and patients
 Validation in the policy forum

Milestone 4 – January 2023

Main result – final recommendations

Upgrading the document
 Inclusion in the DQAF

Identification of relevant standards and data models for semantic harmonization

